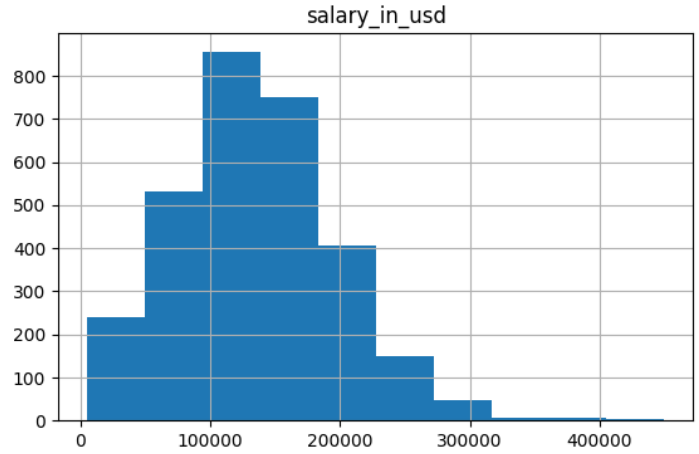
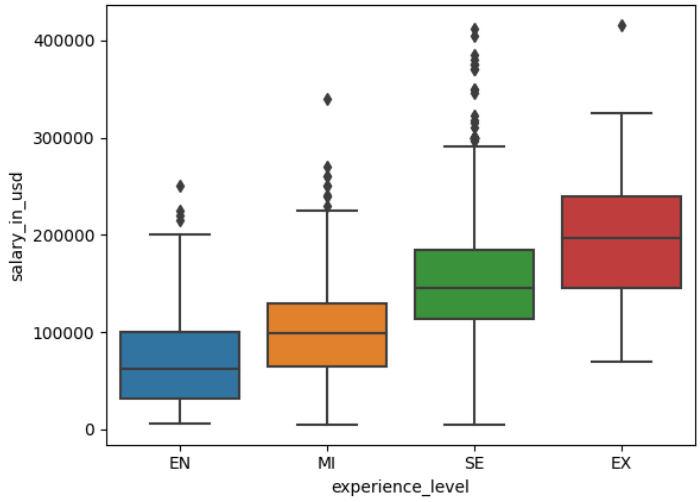
What is the appropriate salary for a Data Science and Machine Learning employee based on their experience, work modality, and location?

To get a first impression of the salary distribution, I plotted a histogram of Salary in USD to have uniformity between currencies:

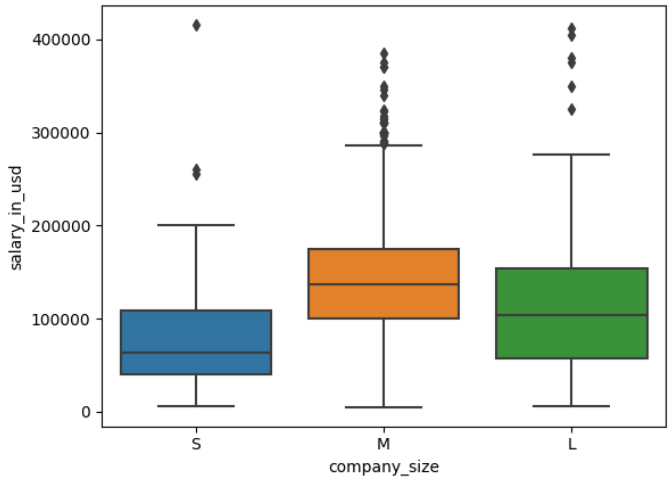


Three of the highest salaries appeared to be error outliers, since their experience level was Middle, and they were well above Senior and Executive salaries. So I dropped these 3 entries.

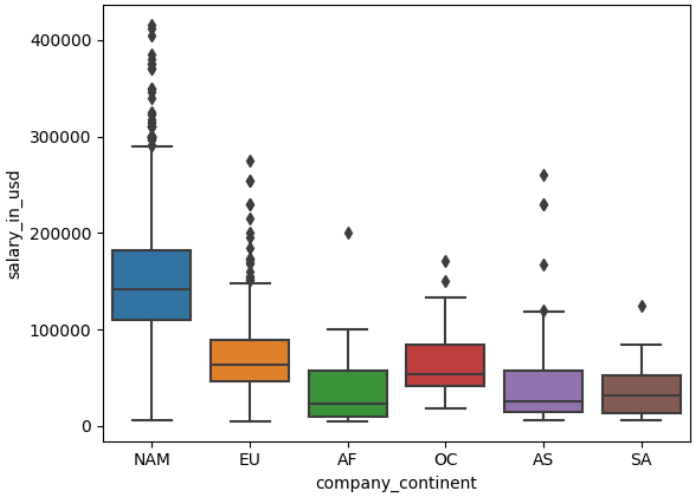


The job_title column had 89 unique entries. Since there are so many variations in the job titles and these can be misleading I decided to drop the job title and focus on experience level, which the graph above shows to be a more intuitive relationship.

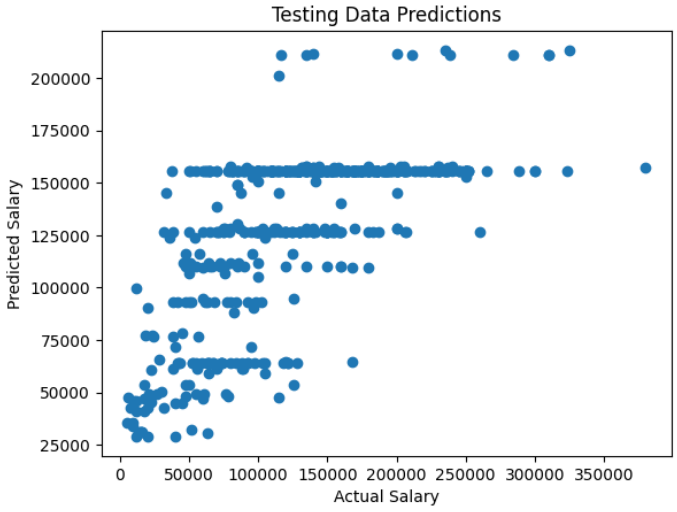
While it might seem like salary would be directly correlated with company size, we can see that on average medium size companies pay a higher salary, which on further thought makes sense, since they have more employees to pay.



The data on the location columns was reported by country, which produced many different unique entries, some of which had few reported salaries. Therefore I grouped the countries into continents to get a more robust distribution between locations.



I trained a Linear Regression Model, Ridge Model, and Random Forest Classifier Model on our data:



They all produced very similar results with an R-Score of around 0.40, which is not a very accurate model. Our predictions seem to be off by around $37,000 on average.

We can see many of the predictions are boxed in clusters as opposed to the actual salaries. The outliers seem to be affecting the model negatively, as our predictions cap at around $225,000, while the actual salaries go all the way up to $400,000

To get more accurate results, we might try getting more entries on our data and make sure the information is accurate, since these are all self-reported, and there could be mistakes when reporting the salary in USD.